

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (Error! Hyperlink reference not valid.) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	The effect of alcohol consumption on hemoglobin level among non-pregnant reproductive age women in Ethiopia: a cross sectional secondary data analysis of the 2016 Ethiopian Demographic Health Survey
AUTHORS	Diress, Gedefaw; Endalifer, Melese Linger

VERSION 1 – REVIEW

REVIEWER	Belay, Tafere Central Washington University, Health Sciences
REVIEW RETURNED	09-Dec-2020

GENERAL COMMENTS	<p>An interesting study.</p> <p>Title: the title is a kind of misleading because it looks like it is encouraging alcohol consumption. I would include 'local drinks' into the title. For example - 'The association between alcohol consumption from local drinks and hemoglobin level among.....'</p> <p>Discussion:</p> <p>as you have mentioned in your discussion the local drinks (considered alcohol containing drinks) are made up of grains and gains are sources of iron however grains also contain phytate; a compound which is a major inhibitor of iron uptake. Now, I would suggest you to include the benefit of fermentation in breaking down phytate and enhances iron absorption. As you may know all of the local drinks in Ethiopia are made up of fermented products. Please include references in your discussion (example: from Gibson study).</p> <p>References</p> <p>Some references are too old and irrelevant ex. reference 9, 20, 22, etc.</p>
-------------------------	---

REVIEWER	Kibret, Kelemu Wollega University, Public Health
REVIEW RETURNED	26-Jan-2021

GENERAL COMMENTS	<p>Comments to the Authors</p> <p>Method</p> <p>1. The method lacks a brief description of the study design, the study population</p>
-------------------------	---

	<p>2. Better to describe how the samples were determined and selected in EDHS</p> <p>3.“ Media exposure was classified based on response to how often respondents read a newspaper, listened to the radio, or watched television. Those who responded at least once a week to any of these sources were considered to have access to media”. How valid and reliable considering have access media based on this classification? It needs to be clarified and presented scientifically.</p> <p>4. What is the rationale or your basis for your categorization of some variables such as region (major, developing, and major city administrative); occupation (working or not working outside the home at the time of the survey); marital status (married and not married) and wealth index (poor, middle, and rich)</p> <p>5. “Access to care is defined as having visited a healthcare facility within 12 months of the survey”. Is it standard definition? If so, put the reference.</p> <p>6. “ All statistical techniques used a complex sampling design applied in the 2016 EDHS used a two-stage stratified sampling technique” this statement is not clear. The author should describe the specific statistical technique (s) used to address the complex and hierarchical nature of the EDHS data. As the EDHS collected through a hierarchical multistage sampling, the data is a correlated and the analysis methods should handle the correlated data e.g. multilevel analysis.</p> <p>7. Which statistical package was used for analysis?</p> <p>8. Why the authors used logistics regression as the issue already addressed by linear regression. Is there any reason for applying logistics regression as the objective of this research is to assess the association between alcohol consumption and haemoglobin level?</p> <p>9. How the linear regression model applied for the categorical variables? The linear regression and ANOVA need more detail description rather than simply naming only.</p> <p>10. There are no statements concerning ethical considerations</p> <p>Results</p> <p>11. The results of multivariable linear regression not adequately described and presented. How the variables were selected for multivariable linear regression? What is the R² value for the final model? Where is the multi-collinearity test result?</p> <p>12. Even if the authors' main focus is alcohol drinking, the results of all exposure variables should be presented in table 2 and table 3.</p> <p>13. I would recommend doing the post hoc ANOVA test.</p> <p>Conclusion</p> <p>14. “The prevalence of anaemia among non-pregnant women was relatively high” ... relative to what?</p> <p>15. How would it be practical and ethical recommending experimental study for this issue?</p>
--	--

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

Title

Comment 1: *The title is a kind of misleading because it looks like it is encouraging alcohol consumption. I would include 'local drinks' into the title. For example - 'The association between alcohol consumption from local drinks and hemoglobin level among.....'*

Response1: Thank you very much. We agree with this comment but the types of alcohol was not mentioned in 2016 Ethiopian Demographic Health Survey. To make the title clear we modified as follow: The effect of alcohol consumption on hemoglobin level among non-pregnant reproductive age women in Ethiopia: a cross sectional secondary data analysis of the 2016 Ethiopian Demographic Health Survey

Discussion

Comment 2: *as you have mentioned in your discussion the local drinks (considered alcohol containing drinks) are made up of grains and gains are sources of iron however grains also contain phytate; a compound which is a major inhibitor of iron uptake. Now, I would suggest you to include the benefit of fermentation in breaking down phytate and enhances iron absorption. As you may know all of the local drinks in Ethiopia are made up of fermented products. Please include references in your discussion (example: from Gibson study).*

Response2: Thank you very much for your valuable comment. We strongly agree and corrected. This change included in the revised manuscript on discussion section on page number 15.

References

Comment 3: *Some references are too old and irrelevant ex. reference 9, 20, 22, etc.*

Response 3: We accepted the comment. We removed the outdated references.

Reviewer 2

Methods

Comment 1: The method lacks a brief description of the study design, the study population

Response1: We included the detailed description of the methods used in 2016 EDHS in the revised manuscript. Regarding to study population, as described in the original manuscript, all non-pregnant reproductive age women and women who had no hemoglobin measurement were the study population.

Comment 2: *Better to describe how the samples were determined and selected in EDHS*

Response2: we accept the comment and it is described in the revised manuscript.

Comment 3: *“Media exposure was classified based on response to how often respondents read a newspaper, listened to the radio, or watched television. Those who responded at least once a week to any of these sources were considered to have access to media”. How valid and reliable considering have access media based on this classification? It needs to be clarified and presented scientifically.*

Response 3: In this study, we classified the level of exposure to mass media based the 2016 Ethiopian Demographic and Health Survey classification. In 2016, EDHS participants were asked how often they read a newspaper, listened to the radio, or watched television. Those who responded at least once a week were considered to be regularly exposed to that form of media.

Comment 4: *What is the rationale or your basis for your categorization of some variables such as region (major, developing, and major city administrative); occupation (working or not working outside the home at the time of the survey); marital status (married and not married) and wealth index (poor, middle, and rich)*

Response 4: Ethiopia is administratively divided into 9 regional states and 2 city administration. However, the Ethiopian government classified 9 regional states in to two categories based on economic growth and availability of infrastructure. Oromia, Amhara Tigray and Southern Nation, Nationality and People (SNNP) were classified as developed regional states. Whereas, Afar, Somali, Gambella, Harari and Benishangul-Gumuz were considered as developing regional state. Addis Ababa and Dire Dawa were the two city administration.

Some variables like occupation, wealth index and marital status were categorized to make them suitable for analysis and classified based on previous studies done in Ethiopia. The references are included in the revised manuscript. The 2016 EDHS datasets contained wealth index that was created using principal components analysis (PCA).

Comment 5: *“Access to care is defined as having visited a healthcare facility within 12 months of the survey”. Is it standard definition? If so, put the reference*

Response 5: It is not standard definition but we put EDHS definition of access to care.

Comment 6: *“All statistical techniques used a complex sampling design applied in the 2016 EDHS used a two-stage stratified sampling technique” this statement is not clear. The author should describe the specific statistical technique (s) used to address the complex and hierarchical nature of the EDHS data. As the EDHS collected through a hierarchical multistage sampling, the data is a correlated and the analysis methods should handle the correlated data e.g. multilevel analysis.*

Response 6: We agree with the comment. As you know, in complex sample design, when standard errors, confidence intervals or significance testing is required, it is necessary to consider three pieces of information: the primary sampling unit or cluster variable, the stratification variable, and the weight variable. If we assume simple random sampling and independence of observations, the standard errors will likely be underestimated and the resulting significance tests invalid. This is included in the revised manuscript. Even though the data have hierarchical nature we didn't consider cluster level factors using multilevel analysis because our aim is only to assess the effect of alcohol drinking on hemoglobin level.

Comment 7: *Which statistical package was used for analysis?*

Response 7: The analysis was carried out using Statistical Package for Social Science (SPSS) version 25. The change is included in the revised manuscript.

Comment 8: *Why the authors used logistics regression as the issue already addressed by linear regression. Is there any reason for applying logistics regression as the objective of this research is to assess the association between alcohol consumption and haemoglobin level? .*

Response 8: In addition to linear regression we used logistic regression because linear regression is only strong enough for prediction of the response variable (in our case hemoglobin level). However, logistic regression is strong enough for classification purposes of the response variable when we assess the association between the response variable and the independent variable. Therefore, logistic regression is used for solving classification problems of the dependent variables. Besides, linear regression require linear relationship between dependent variable and independent variable. To detect nonlinear relationship between alcohol drinking and anemia we used logistic regression.

Comment 9: *How the linear regression model applied for the categorical variables? The linear regression and ANOVA need more detail description rather than simply naming only.*

Response 9: linear regression model is applied when the dependent variable is continuous type of variable. In our study, hemoglobin level is continuous type of variable. For logistic regression the hemoglobin level is dichotomized to anemic and non- anemic. Based on your comment, we also added some description about linear regression and ANOVA in the revised manuscript..

Comment 10: *There are no statements concerning ethical considerations*

Response 10: The study is based on secondary data analysis of EDHS. Since secondary data from DHS provided by Marco International was used, there was no ethical approval required. It should however, be noted that Macro International obtained informed consent from individuals who participated in the DHS surveys. We accessed the data from MEASURE DHS database at <http://dhsprogram.com/data/>.

According to BMJ Open journal, ethical consideration should be described under declaration section.

Therefore, ethical approval is '**not applicable**' for this study which is stated under declaration section next to references.

Result

Comment 11: *The results of multivariable linear regression not adequately described and presented. How the variables were selected for multivariable linear regression? What is the R2 value for the final model? Where is the multi-collinearity test result?*

Response 11: Thank you very much for your essential comments. All variables with p-value less than 0.2 at bivariate analysis (simple linear regression) were entered into final regression model. Based on your comment, we mentioned this in the method and result sections in the revised manuscript.

The value of R^2 for the final model is 0.147. This indicates that 14.7 % of the variance in hemoglobin level can be predicted from the variables alcohol consumption, age, educational status, marital status, occupation, wealth index, media exposure, BMI, contraceptive, number of birth in the last 5 years, access to care and source of drinking water. This change is included in the revised manuscript.

Multicollinearity was assessed using VIF values. The result showed that all VIF values were below 10, shows absence of multicollinearity between predictor variables. The finding is showed in the next table.

Coefficients ^a								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	11.859	.077		154.275	.000		
	age catagorized	-.033	.022	-.015	-1.497	.134	.700	1.428
	BMI catagorized	.187	.027	.062	7.043	.000	.912	1.097
	occupation	.236	.031	.066	7.629	.000	.927	1.078
	marital status	-.290	.040	-.073	-7.280	.000	.686	1.457
	wealth index	.286	.018	.148	15.742	.000	.784	1.275
	source of water	-.002	.018	-.001	-.107	.915	.878	1.139
	alcohol exposure	.263	.020	.113	13.172	.000	.951	1.051

a. Dependent Variable: hemoglobin

Comment 12: *Even if the authors' main focus is alcohol drinking, the results of all exposure variables should be presented in table 2 and table 3.*

Response 12: Yes you are right. However, as you mentioned, our main objective is only to assess the effect of alcohol drinking on hemoglobin level (anemia). To be focused and in line with our title and objective, we preferred to present only the main exposure variable (alcohol drinking) in the regression table. To be sure you can checked the adjusted variable in the next final regression model (at footnote).

Odds Ratios 2^a

			Odds Ratio	95% Confidence Interval	
binary anemia				Lower	Upper
alcohol exposure	less than one per week vs. No	Has anemia	.540	.311	.939
	at least once a week vs. No	Has anemia	.498	.283	.877
	almost every day vs. No	Has anemia	.415	.214	.805

Dependent Variable: binary anemia (reference category = No anemia)

Model: (Intercept), exposure, V025, S1107C, BMlcat, cuurentFP, birthlast5, wealthinexcat, catage, occupatcat, maritalstacat, watersource

a. Factors and covariates used in the computation are fixed at the following values: alcohol exposure=almost every day; Type of place of residence=Rural; Have you ever taken a drink that contains alcohol=Yes; BMI catagorized= ≥ 25 ; current contraceptive=non hormonal and traditional; birth in last 5 years= ≥ 3 birth; wealth index=rich; age catagorized=35-49; occupation=Working; marital status=Married; source of water=Not improved

Comment 13: *I would recommend doing the post hoc ANOVA test.*

Response 13: Thank you very much for pointing out this. Our data met the assumption of homogeneity of variances. Hence we performed Tukey's honestly significant difference (HSD) post hoc test. The post hoc test result showed that there is significant difference between nondrinkers and all the remaining three groups (less than once a week, at least once a week, and almost every day) (p -value <0.001). However, there is no significant difference between a group who drink less than once a week and drink at least once a week (p -value=0.912) and there is significant difference between participants who drink alcohol less than a week and who drink almost every day (p -value=0.959). Similarly, there is no significant difference between the last two groups (at least once a week, and almost every day) (p -value=0.999).

Post Hoc Tests

Multiple Comparisons						
Dependent Variable: hemoglobin						
Tukey HSD						
(I) alcohol exposure	(J) alcohol exposure	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
No	less than one per week	-.60348 [*]	.03836	.000	-.7020	-.5049
	at least once a week	-.56535 [*]	.05055	.000	-.6952	-.4355
	almost every day	-.54772 [*]	.10752	.000	-.8240	-.2715
less than one per week	No	.60348 [*]	.03836	.000	.5049	.7020
	at least once a week	.03813	.05776	.912	-.1103	.1865
	almost every day	.05577	.11109	.959	-.2297	.3412
at least once a week	No	.56535 [*]	.05055	.000	.4355	.6952
	less than one per week	-.03813	.05776	.912	-.1865	.1103
	almost every day	.01764	.11586	.999	-.2801	.3153
almost every day	No	.54772 [*]	.10752	.000	.2715	.8240
	less than one per week	-.05577	.11109	.959	-.3412	.2297
	at least once a week	-.01764	.11586	.999	-.3153	.2801

*. The mean difference is significant at the 0.05 level.

Conclusion

Comment 14: *"The prevalence of anaemia among non-pregnant women was relatively high" ... relative to what?*

Response 14: we tried to compare the current finding to previous studies. However, to make clear and to avoid confusion we have modified the sentence in the revised manuscript.

Comment 15: *How would it be practical and ethical recommending experimental study for this issue?*

Response 15: Thank you very much. Yes you are right. It is unethical to do experimental studies for known harmful exposure. We modified the recommendation.